

#12
BP 1635
6-21-01

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/530,663A

DATE: 05/30/2001

TIME: 12:35:13

Input Set : A:\B0045591.txt

Output Set: C:\CRF3\05302001\I530663A.raw

RECEIVED

JUN 21 2001

TECH CENTER 1600/2900

ENTERED

4 <110> APPLICANT: Chiang, Vincent Lee C.
 5 Tsai, Chung-Jui
 6 Hu, Wen-Jing
 8 <120> TITLE OF INVENTION: GENETIC ENGINEERING OF LIGNIN BIOSYNTHESIS IN PLANTS
 10 <130> FILE REFERENCE: 66040-9651
 12 <140> CURRENT APPLICATION NUMBER: 09/530,663A
 13 <141> CURRENT FILING DATE: 2000-07-11
 15 <150> PRIOR APPLICATION NUMBER: PCT/US98/24138
 16 <151> PRIOR FILING DATE: 1998-11-12
 18 <150> PRIOR APPLICATION NUMBER: 08/969,046
 19 <151> PRIOR FILING DATE: 1997-11-12
 21 <160> NUMBER OF SEQ ID NOS: 17
 23 <170> SOFTWARE: Word 97 (DOS text format)
 25 <210> SEQ ID NO: 1
 26 <211> LENGTH: 1915
 27 <212> TYPE: DNA
 28 <213> ORGANISM: Populus tremuloides Michx. (aspen)
 30 <220> FEATURE:
 31 <221> NAME/KEY: CDS
 32 <222> LOCATION: (83)...(1687)
 34 <400> SEQUENCE: 1

```

35 ccctcgcgaa actccgaaaa cagagagcac ctaaaactca ccatctctcc ctctgcatct      60
36 ttagcccgca atggacgcca ca atg aat cca caa gaa ttc atc ttt cgc tca      112
37                               Met Asn Pro Gln Glu Phe Ile Phe Arg Ser
38                               1                               5                               10
40 aaa tta cca gac atc tac atc ccg aaa aac ctt ccc ctg cat tca tac      160
41 Lys Leu Pro Asp Ile Tyr Ile Pro Lys Asn Leu Pro Leu His Ser Tyr
42                               15                               20                               25
44 gtt ctt gag aac ttg tct aaa cat tca tca aaa cct tgc ctg ata aat      208
45 Val Leu Glu Asn Leu Ser Lys His Ser Ser Lys Pro Cys Leu Ile Asn
46                               30                               35                               40
48 ggc gcg aat gga gat gtc tac acc tat gct gat gtt gag ctc aca gca      256
49 Gly Ala Asn Gly Asp Val Tyr Thr Tyr Ala Asp Val Glu Leu Thr Ala
50                               45                               50                               55
52 aga aga gtt gct tct ggt ctg aac aag att ggt att caa caa ggt gac      304
53 Arg Arg Val Ala Ser Gly Leu Asn Lys Ile Gly Ile Gln Gln Gly Asp
54                               60                               65                               70
56 gtg atc atg ctc ttc cta cca agt tca cct gaa ttc gtg ctt gct ttc      352
57 Val Ile Met Leu Phe Leu Pro Ser Ser Pro Glu Phe Val Leu Ala Phe
58                               75                               80                               85                               90
60 cta ggc gct tca cac aga ggt gcc atg atc act gct gcc aat cct ttc      400
61 Leu Gly Ala Ser His Arg Gly Ala Met Ile Thr Ala Ala Asn Pro Phe
62                               95                               100                               105
64 tcc acc cct gca gag cta gca aaa cat gcc aag gcc tcg aga gca aag      448
65 Ser Thr Pro Ala Glu Leu Ala Lys His Ala Lys Ala Ser Arg Ala Lys
66                               110                               115                               120
68 ctt ctg ata aca cag gct tgt tac tac gag aag gtt aaa gat ttt gcc      496

```

RAW SEQUENCE LISTING

DATE: 05/30/2001

PATENT APPLICATION: US/09/530,663A

TIME: 12:35:13

Input Set : A:\B0045591.txt

Output Set: C:\CRF3\05302001\I530663A.raw

69	Leu	Leu	Ile	Thr	Gln	Ala	Cys	Tyr	Tyr	Glu	Lys	Val	Lys	Asp	Phe	Ala	
70			125					130					135				
72	cga	gaa	agt	gat	gtt	aag	gtc	atg	tgc	gtg	gac	tct	gcc	ccg	gac	ggg	544
73	Arg	Glu	Ser	Asp	Val	Lys	Val	Met	Cys	Val	Asp	Ser	Ala	Pro	Asp	Gly	
74		140					145					150					
76	gct	tca	ctt	ttc	aga	gct	cac	aca	cag	gca	gac	gaa	aat	gaa	gtg	cct	592
77	Ala	Ser	Leu	Phe	Arg	Ala	His	Thr	Gln	Ala	Asp	Glu	Asn	Glu	Val	Pro	
78	155					160					165				170		
80	cag	gtc	gac	att	agt	cct	gat	gat	gtc	gta	gca	ttg	cct	tat	tca	tca	640
81	Gln	Val	Asp	Ile	Ser	Pro	Asp	Asp	Val	Val	Ala	Leu	Pro	Tyr	Ser	Ser	
82				175					180				185				
84	ggg	act	aca	ggg	ttg	cca	aaa	ggg	gtc	atg	tta	acg	cac	aaa	ggg	cta	688
85	Gly	Thr	Thr	Gly	Leu	Pro	Lys	Gly	Val	Met	Leu	Thr	His	Lys	Gly	Leu	
86			190						195				200				
88	ata	acc	agt	gtg	gct	caa	cag	gta	gat	gga	gac	aat	cct	aac	ctg	tat	736
89	Ile	Thr	Ser	Val	Ala	Gln	Gln	Val	Asp	Gly	Asp	Asn	Pro	Asn	Leu	Tyr	
90		205					210					215					
92	ttt	cac	agt	gaa	gat	gtg	att	ctg	tgt	gtg	ctt	cct	atg	ttc	cat	atc	784
93	Phe	His	Ser	Glu	Asp	Val	Ile	Leu	Cys	Val	Leu	Pro	Met	Phe	His	Ile	
94		220				225					230						
96	tat	gct	ctg	aat	tca	atg	atg	ctc	tgt	ggg	ctg	aga	ggt	ggg	gcc	tcg	832
97	Tyr	Ala	Leu	Asn	Ser	Met	Met	Leu	Cys	Gly	Leu	Arg	Val	Gly	Ala	Ser	
98	235			240					245				250				
100	att	ttg	ata	atg	cca	aag	ttt	gag	att	ggg	tct	ttg	ctg	gga	ttg	att	880
101	Ile	Leu	Ile	Met	Pro	Lys	Phe	Glu	Ile	Gly	Ser	Leu	Leu	Gly	Leu	Ile	
102			255					260				265					
104	gag	aag	tac	aag	gta	tct	ata	gca	cca	ggt	ggt	cca	cct	gtg	atg	atg	928
105	Glu	Lys	Tyr	Lys	Val	Ser	Ile	Ala	Pro	Val	Val	Pro	Pro	Val	Met	Met	
106			270					275				280					
108	gca	att	gct	aag	tca	cct	gat	ctt	gac	aag	cat	gac	ctg	tct	tct	ttg	976
109	Ala	Ile	Ala	Lys	Ser	Pro	Asp	Leu	Asp	Lys	His	Asp	Leu	Ser	Ser	Leu	
110		285					290					295					
112	agg	atg	ata	aaa	tct	gga	ggg	gct	cca	ttg	ggc	aag	gaa	ctt	gaa	gat	1024
113	Arg	Met	Ile	Lys	Ser	Gly	Gly	Ala	Pro	Leu	Gly	Lys	Glu	Leu	Glu	Asp	
114		300				305					310						
116	act	gtc	aga	gct	aag	ttt	cct	cag	gct	aga	ctt	ggg	cag	gga	tat	gga	1072
117	Thr	Val	Arg	Ala	Lys	Phe	Pro	Gln	Ala	Arg	Leu	Gly	Gln	Gly	Tyr	Gly	
118	315			320						325			330				
120	atg	acc	gag	gca	gga	cct	ggt	cta	gca	atg	tgc	ttg	gca	ttt	gcc	aag	1120
121	Met	Thr	Glu	Ala	Gly	Pro	Val	Leu	Ala	Met	Cys	Leu	Ala	Phe	Ala	Lys	
122			335					340				345					
124	gaa	cca	ttc	gac	ata	aaa	cca	ggg	gca	tgt	gga	act	gta	gtc	agg	aat	1168
125	Glu	Pro	Phe	Asp	Ile	Lys	Pro	Gly	Ala	Cys	Gly	Thr	Val	Val	Arg	Asn	
126		350						355				360					
128	gca	gag	atg	aag	att	gtt	gac	cca	gaa	aca	ggg	gtc	tct	cta	ccg	agg	1216
129	Ala	Glu	Met	Lys	Ile	Val	Asp	Pro	Glu	Thr	Gly	Val	Ser	Leu	Pro	Arg	
130		365					370					375					
132	aac	cag	cct	ggg	gag	atc	tgc	atc	cgg	ggg	gat	cag	atc	atg	aaa	gga	1264
133	Asn	Gln	Pro	Gly	Glu	Ile	Cys	Ile	Arg	Gly	Asp	Gln	Ile	Met	Lys	Gly	

RAW SEQUENCE LISTING

DATE: 05/30/2001

PATENT APPLICATION: US/09/530,663A

TIME: 12:35:13

Input Set : A:\B0045591.txt

Output Set: C:\CRF3\05302001\I530663A.raw

```

134      380      385      390
136 tat ctt aat gac ccc gag gca acc tca aga aca ata gac aaa gaa gga      1312
137 Tyr Leu Asn Asp Pro Glu Ala Thr Ser Arg Thr Ile Asp Lys Glu Gly
138 395      400      405      410
140 tgg ctg cac aca ggc gat atc ggc tac att gat gat gat gag ctt      1360
141 Trp Leu His Thr Gly Asp Ile Gly Tyr Ile Asp Asp Asp Asp Glu Leu
142      415      420      425
144 ttc atc gtt gac aga ttg aag gaa ttg atc aag tat aaa ggg ttt cag      1408
145 Phe Ile Val Asp Arg Leu Lys Glu Leu Ile Lys Tyr Lys Gly Phe Gln
146      430      435      440
148 gtt gct cct act gaa ctc gaa gct ttg tta ata gcc cat cca gag ata      1456
149 Val Ala Pro Thr Glu Leu Glu Ala Leu Leu Ile Ala His Pro Glu Ile
150      445      450      455
152 tcc gat gct gct gta gta gga ttg aaa gat gag gat gcg gga gaa gtt      1504
153 Ser Asp Ala Ala Val Val Gly Leu Lys Asp Glu Asp Ala Gly Glu Val
154      460      465      470
156 cct gtt gca ttt gta gtg aaa tca gaa aag tct cag gcc acc gaa gat      1552
157 Pro Val Ala Phe Val Val Lys Ser Glu Lys Ser Gln Ala Thr Glu Asp
158 475      480      485      490
160 gaa att aag cag tat att tca aaa cag gtg atc ttc tac aag aga ata      1600
161 Glu Ile Lys Gln Tyr Ile Ser Lys Gln Val Ile Phe Tyr Lys Arg Ile
162      495      500      505
164 aaa cga gtt ttc ttc att gaa gca att ccc aag gca cca tca ggc aag      1648
165 Lys Arg Val Phe Phe Ile Glu Ala Ile Pro Lys Ala Pro Ser Gly Lys
166      510      515      520
168 atc ctg agg aag aat ctg aaa gag aag ttg cca ggc ata taactgaaga      1697
169 Ile Leu Arg Lys Asn Leu Lys Glu Lys Leu Pro Gly Ile
170      525      530      535
172 tggtactgaa catttaaccc tctgtcttat ttctttaata cttgcgaatc attgtagtgt      1757
173 tgaaccaagc atgcttgaa aagacacgta cccaacgtaa gacagttact gttcctagta      1817
174 tacaagctct ttaatgttcg ttttgaactt gggaaaacat aagttctcct gtcgccatat      1877
175 ggagtaattc aattgaatat tttggtttct ttaatgat      1915
178 <210> SEQ ID NO: 2
179 <211> LENGTH: 535
180 <212> TYPE: PRT
181 <213> ORGANISM: Populus tremuloides Michx. (aspen)
183 <400> SEQUENCE: 2
184 Met Asn Pro Gln Glu Phe Ile Phe Arg Ser Lys Leu Pro Asp Ile Tyr
185 1      5      10      15
186 Ile Pro Lys Asn Leu Pro Leu His Ser Tyr Val Leu Glu Asn Leu Ser
187 20      25      30
188 Lys His Ser Ser Lys Pro Cys Leu Ile Asn Gly Ala Asn Gly Asp Val
189 35      40      45
190 Tyr Thr Tyr Ala Asp Val Glu Leu Thr Ala Arg Arg Val Ala Ser Gly
191 50      55      60
192 Leu Asn Lys Ile Gly Ile Gln Gln Gly Asp Val Ile Met Leu Phe Leu
193 65      70      75      80
194 Pro Ser Ser Pro Glu Phe Val Leu Ala Phe Leu Gly Ala Ser His Arg
195      85      90      95

```

RAW SEQUENCE LISTING

DATE: 05/30/2001

PATENT APPLICATION: US/09/530,663A

TIME: 12:35:13

Input Set : A:\B0045591.txt

Output Set: C:\CRF3\05302001\I530663A.raw

```

196 Gly Ala Met Ile Thr Ala Ala Asn Pro Phe Ser Thr Pro Ala Glu Leu
197           100           105           110
198 Ala Lys His Ala Lys Ala Ser Arg Ala Lys Leu Leu Ile Thr Gln Ala
199           115           120           125
200 Cys Tyr Tyr Glu Lys Val Lys Asp Phe Ala Arg Glu Ser Asp Val Lys
201           130           135           140
202 Val Met Cys Val Asp Ser Ala Pro Asp Gly Ala Ser Leu Phe Arg Ala
203           145           150           155           160
204 His Thr Gln Ala Asp Glu Asn Glu Val Pro Gln Val Asp Ile Ser Pro
205           165           170           175
206 Asp Asp Val Val Ala Leu Pro Tyr Ser Ser Gly Thr Thr Gly Leu Pro
207           180           185           190
208 Lys Gly Val Met Leu Thr His Lys Gly Leu Ile Thr Ser Val Ala Gln
209           195           200           205
210 Gln Val Asp Gly Asp Asn Pro Asn Leu Tyr Phe His Ser Glu Asp Val
211           210           215           220
212 Ile Leu Cys Val Leu Pro Met Phe His Ile Tyr Ala Leu Asn Ser Met
213           225           230           235           240
214 Met Leu Cys Gly Leu Arg Val Gly Ala Ser Ile Leu Ile Met Pro Lys
215           245           250           255
216 Phe Glu Ile Gly Ser Leu Leu Gly Leu Ile Glu Lys Tyr Lys Val Ser
217           260           265           270
218 Ile Ala Pro Val Val Pro Pro Val Met Met Ala Ile Ala Lys Ser Pro
219           275           280           285
220 Asp Leu Asp Lys His Asp Leu Ser Ser Leu Arg Met Ile Lys Ser Gly
221           290           295           300
222 Gly Ala Pro Leu Gly Lys Glu Leu Glu Asp Thr Val Arg Ala Lys Phe
223           305           310           315           320
224 Pro Gln Ala Arg Leu Gly Gln Gly Tyr Gly Met Thr Glu Ala Gly Pro
225           325           330           335
226 Val Leu Ala Met Cys Leu Ala Phe Ala Lys Glu Pro Phe Asp Ile Lys
227           340           345           350
228 Pro Gly Ala Cys Gly Thr Val Val Arg Asn Ala Glu Met Lys Ile Val
229           355           360           365
230 Asp Pro Glu Thr Gly Val Ser Leu Pro Arg Asn Gln Pro Gly Glu Ile
231           370           375           380
232 Cys Ile Arg Gly Asp Gln Ile Met Lys Gly Tyr Leu Asn Asp Pro Glu
233           385           390           395           400
234 Ala Thr Ser Arg Thr Ile Asp Lys Glu Gly Trp Leu His Thr Gly Asp
235           405           410           415
236 Ile Gly Tyr Ile Asp Asp Asp Asp Glu Leu Phe Ile Val Asp Arg Leu
237           420           425           430
238 Lys Glu Leu Ile Lys Tyr Lys Gly Phe Gln Val Ala Pro Thr Glu Leu
239           435           440           445
240 Glu Ala Leu Leu Ile Ala His Pro Glu Ile Ser Asp Ala Ala Val Val
241           450           455           460
242 Gly Leu Lys Asp Glu Asp Ala Gly Glu Val Pro Val Ala Phe Val Val
243           465           470           475           480
244 Lys Ser Glu Lys Ser Gln Ala Thr Glu Asp Glu Ile Lys Gln Tyr Ile

```

RAW SEQUENCE LISTING

DATE: 05/30/2001

PATENT APPLICATION: US/09/530,663A

TIME: 12:35:13

Input Set : A:\B0045591.txt

Output Set: C:\CRF3\05302001\I530663A.raw

```

245          485          490          495
246 Ser Lys Gln Val Ile Phe Tyr Lys Arg Ile Lys Arg Val Phe Phe Ile
247          500          505          510
248 Glu Ala Ile Pro Lys Ala Pro Ser Gly Lys Ile Leu Arg Lys Asn Leu
249          515          520          525
250 Lys Glu Lys Leu Pro Gly Ile
251          530          535
254 <210> SEQ ID NO: 3
255 <211> LENGTH: 1710
256 <212> TYPE: DNA
257 <213> ORGANISM: Populus tremuloides Michx. (aspen)
259 <220> FEATURE:
260 <221> NAME/KEY: CDS
261 <222> LOCATION: (1)...(1710)
263 <400> SEQUENCE: 3
264 atg atg tcc gtg gcc acg gtt gag ccc ccg aaa ccg gaa ctc tcc cct      48
265 Met Met Ser Val Ala Thr Val Glu Pro Pro Lys Pro Glu Leu Ser Pro
266 1          5          10          15
268 cca caa aac caa aac gca cca tcc tct cat gaa act gat cac att ttc      96
269 Pro Gln Asn Gln Asn Ala Pro Ser Ser His Glu Thr Asp His Ile Phe
270          20          25          30
272 aga tca aaa cta cca gac ata acc atc tcg aac gac ctc cct ctg cac      144
273 Arg Ser Lys Leu Pro Asp Ile Thr Ile Ser Asn Asp Leu Pro Leu His
274          35          40          45
276 gca tac tgc ttt gaa aac ctc tct gat ttc tca gat agg cca tgc ttg      192
277 Ala Tyr Cys Phe Glu Asn Leu Ser Asp Phe Ser Asp Arg Pro Cys Leu
278          50          55          60
280 att tca ggt tcc acg gga aaa acc tat tct ttt gcc gaa act cac ctc      240
281 Ile Ser Gly Ser Thr Gly Lys Thr Tyr Ser Phe Ala Glu Thr His Leu
282 65          70          75          80
284 ata tct cgg aag gtc gct gct ggg tta tcc aat ttg ggc atc aag aaa      288
285 Ile Ser Arg Lys Val Ala Ala Gly Leu Ser Asn Leu Gly Ile Lys Lys
286          85          90          95
288 ggc gat gta atc atg acc ctg ctc caa aac tgc cca gaa ttc gtc ttc      336
289 Gly Asp Val Ile Met Thr Leu Leu Gln Asn Cys Pro Glu Phe Val Phe
290          100          105          110
292 tcc ttc atc ggt gct tcc atg att ggt gca gtc atc acc act gcg aac      384
293 Ser Phe Ile Gly Ala Ser Met Ile Gly Ala Val Ile Thr Thr Ala Asn
294          115          120          125
296 cct ttc tac act caa agt gaa ata ttc aag caa ttc tct gct tct cgt      432
297 Pro Phe Tyr Thr Gln Ser Glu Ile Phe Lys Gln Phe Ser Ala Ser Arg
298          130          135          140
300 gcg aaa ctg att atc acc cag tct caa tat gtg aac aag cta gga gat      480
301 Ala Lys Leu Ile Ile Thr Gln Ser Gln Tyr Val Asn Lys Leu Gly Asp
302 145          150          155          160
304 agt gat tgc cat gaa aac aac caa aaa ccg ggg gaa gat ttc ata gta      528
305 Ser Asp Cys His Glu Asn Asn Gln Lys Pro Gly Glu Asp Phe Ile Val
306          165          170          175
308 atc acc att gat gac ccg cca gag aac tgt cta cat ttc aat gtg ctt      576

```

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/530,663A

DATE: 05/30/2001

TIME: 12:35:14

Input Set : A:\B0045591.txt

Output Set: C:\CRF3\05302001\I530663A.raw

L:593 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:9
L:616 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:634 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:11